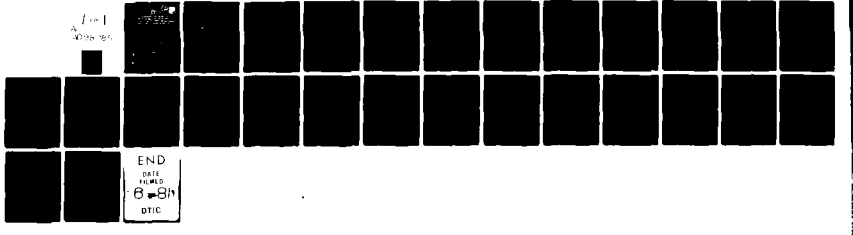
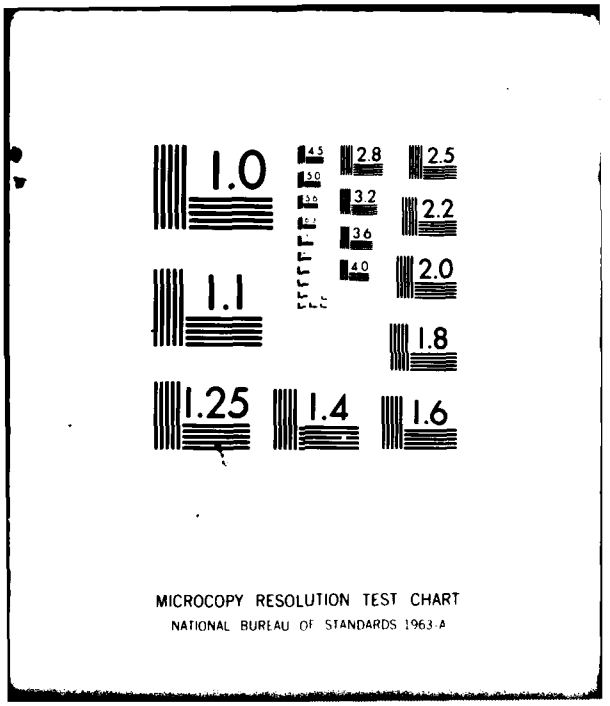
AD-A098 785 SYRACUSE UNIV N Y DEPT OF INDUSTRIAL ENGINEERING AND--ETC F/G 9/2
AN ASSESSMENT PROCEDURE AND A SET OF CRITERIA FOR USE IN THE EV--ETC(U)
FEB 81 R G SARGENT F30602-78-C-0083
UNCLASSIFIED WP-80-016 RADC-TR-80-409 NL



END
DATE FILMED
6-81
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

RADC-TR-80-409
Final Technical Report
February 1981

LEVEL II

12

AD A098785

AN ASSESSMENT PROCEDURE AND A SET OF CRITERIA FOR USE IN THE EVALUATION OF 'COMPUTERIZED MODELS AND COMPUTER-BASED MODELLING TOOLS'

Syracuse University

Robert G. Sargent

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

DTIC
ELECTE
MAY 13 1981
S D
E

ROME AIR DEVELOPMENT CENTER
Air Force Systems Command
Griffiss Air Force Base, New York 13441

DTIC FILE COPY

81 5 13 001

This report has been reviewed by the RADC Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RADC-TR-80-409 has been reviewed and is approved for publication.

APPROVED: Joe L. Thomas

JOE L. THOMAS, 1Lt, USAF
Project Engineer

APPROVED:

JOHN J. MARCINIAK, Colonel, USAF
Chief, Information Sciences Division

FOR THE COMMANDER: John P. Huss

JOHN P. HUSS
Acting Chief, Plans Office

If your address has changed or if you wish to be removed from the RADC mailing list, or if the addressee is no longer employed by your organization, please notify RADC (ISCP), Griffiss AFB NY 13441. This will assist us in maintaining a current mailing list.

Do not return this copy. Retain or destroy.

MISSION
of
Rome Air Development Center

RADC plans and executes research, development, test and selected acquisition programs in support of Command, Control Communications and Intelligence (C^3I) activities. Technical and engineering support within areas of technical competence is provided to ESD Program Offices (POs) and other ESD elements. The principal technical mission areas are communications, electromagnetic guidance and control, surveillance of ground and aerospace objects, intelligence data collection and handling, information system technology, ionospheric propagation, solid state sciences, microwave physics and electronic reliability, maintainability and compatibility.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: RADC-TR-80-409	2. GOVT ACCESSION NO.: AD-A098 785
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): AN ASSESSMENT PROCEDURE AND A SET OF CRITERIA FOR USE IN THE EVALUATION OF COMPUTERIZED MODELS AND COMPUTER-BASED MODELLING TOOLS	5. TYPE OF REPORT & PERIOD COVERED: Final Technical Report
	6. PERFORMING ORG. REPORT NUMBER: WP-80-016
7. AUTHOR(s): Robert G. Sargent	8. CONTRACT OR GRANT NUMBER(s): F30602-78-C-0083
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Syracuse University, Dept of Industrial Engr and Operations Research, Syracuse NY 13210	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 62702F, 55812IP1
11. CONTROLLING OFFICE NAME AND ADDRESS: Rome Air Development Center (ISCP), Griffiss AFB NY 13441	12. REPORT DATE: February 1981
	13. NUMBER OF PAGES: 30
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Same	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE: N/A

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Same

18. SUPPLEMENTARY NOTES

RADC Project Engineer: Joe L. Thomas, 1Lt, USAF (ISCP)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Assessment
Verification
Model Evaluation
Computerized Models
Computer-based Modelling Tools
Validation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper contains a flexible and adaptive assessment procedure and a set of assessment criteria for use in having assessments conducted of general purpose computerized models and computer-based modelling tools. Basically, the procedure requires the use of a third party to make subjective judgments on a set of criteria specified by the assessment sponsor and to make an overall recommendation on the use of the model or modelling tool.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

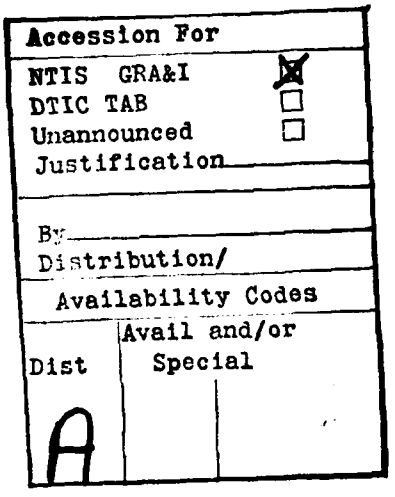
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

EVALUATION

This final technical report presents the work performed by the Department of Industrial Engineering & Operations Research of Syracuse University under contract F30602-78-C00083. The significance of this work in relation to Air Force Technical Objectives is that it presents an assessment approach and a set of criteria for use in performing an evaluation of general purpose computerized models and computer-based modelling tools after they have been developed by the model developers and prior to their use by model users. The set of criteria presented on this work may also be useful in helping to specify to model developers what is expected of them by model sponsors.

Joe L. Thomas

JOE L. THOMAS, 1Lt, USAF
Project Engineer



INTRODUCTION

With the advent of the digital computer, computerized models came into existence. Initially, these models were simple, were developed by the model users themselves, and were used by researchers. Today, computerized models are simple to complex, are used for a wide variety of purposes and applications covering the spectrum from their use in research to describe and understand phenomena to their results being used routinely in day-to-day decision making, and are frequently developed by model developers (builders) who have no direct interaction with the model users or the users of the model's results. Because of this later development, there now exists the problem of ensuring the users of computerized models and their results that they can be used with confidence.

The modelling process, as described in texts and other publications (e.g., [8, 14, 15]), includes the steps of model verification and validation, i.e., ensuring that the model behaves as intended and that there is an adequate agreement between the model and the system (entity) being modelled for the intended usage of the model. Thus one might believe that decision makers and model users should have confidence in computerized models and their results. However, model users and decision makers usually do not have confidence in a model and its results until they have been convinced that it and its results can be trusted. This usually requires, at a minimum, that a model be thoroughly verified and validated and this process be

adequately documented for the decision makers and model users. Unfortunately, as has been pointed out by Gass [3] and others, model developers frequently do not provide adequate documentation on computerized models for a variety of reasons. Thus, even if model developers have sufficiently verified and validated a model, it is usually not adequately documented. Because of this lack of adequate documentation, as well as the desire of the users of computerized models and their results to ensure that computerized models have been properly developed, a new field called model assessment or model evaluation has evolved.

In general, model assessment is concerned with determining the usability of a computerized model. This typically includes at least determining that a computerized model has been properly developed and correctly implemented on the computer, has the desired accuracy necessary for the model's intended use, and has sufficient documentation. As suggested in [6], the objectives of model assessment can be quite varied and includes (1) determining the validity of some specific past applications, (2) determining the validity of some specific future application, (3) determining the areas or problems which a computerized model can contribute to, (4) suggesting improvements in a model, (5) increasing credibility among model users, and (6) testing the usability of a model by others than the model developers. Various procedures and criteria have been used and suggested for computerized model assessment. Gass gives a review of them in [3]. Usually a third or independent group or

party, i.e., a group that is neither the model developer nor the model users, perform or is recommended to perform the model assessment using some procedure or criteria. Unfortunately, however, many assessments are performed without using any specified procedure, criteria, or guidelines.

Research has recently begun in this new field of model assessment which should lead to a clear set of alternative assessment objectives, an adequate set of procedures and criteria to be used in model assessment, and a set of terminology that will help eliminate the current inconsistencies in terminology. Some of the research efforts are general [1, 10, 13], e.g., U.S. General Accounting Office's Exposure Draft on Guidelines for Model Evaluation [5, 15]; however, most of the research is being directed towards how to assess energy models (e.g., [3, 4, 7, 9]). In [3], Gass gives a review of the current research and issues in model assessment.

It is the primary purpose of this paper to present an assessment approach and a set of criteria to be used in performing an evaluation of general purpose computerized models and computer-based modelling tools after they have been developed by the model developers and prior to their use by model users, e.g., system designers. Specifically, they were developed for assessing or evaluating general purpose computerized models and modelling tools for use in the design and evaluation of computer and information systems and subsystems, e.g., computerized models of distributed systems and computer-based modelling tools for designing data base systems. The proce-

dure and criteria presented are flexible and can be easily used and adapted for a wide variety of objectives and applications. They are partially based on previous work, in particular [2, 3, 15].

The remainder of this paper is divided into three sections. The next section presents the assessment procedure, the following section the assessment criteria, and the last section the conclusions.

ASSESSMENT PROCEDURE

The procedure suggested for assessing or evaluating general purpose computerized models and computer-based modelling tools is to use a third or independent party to perform the assessment using a set of criteria. The assesor is to make subjective ratings on how well each criterion is satisfied by selecting a numerical value between zero (0) and ten (10), with zero being the lowest and ten the highest, for each criterion; make an overall recommendation on the use of the computerized model or computer-based modelling tool; and make suggestions for improvements. The set of criteria to be used by the assessor is to be specified by the sponsor of the assessment.

It is suggested that the set of criteria to be used in an assessment be divided into categories by the assessment sponsor. This author believes that the categories of Documentation, Software, Validation, and Overall are appropriate for most assessments of computerized models and computer-based modelling tools. The suggested criteria for each of these categories are listed in Table I and are discussed in the next section. Different categories and criteria can be chosen by the assessment sponsor in using this assessment procedure. Also, if the assessment sponsor desires, they can require the assessor to make a subjective rating between 0 and 10 on how well each category is satisfied.

Regarding the suggestions for improvements, the assessment sponsor can request that they be given by criterion, category, or

TABLE 1

Categories and Their Criteria

I. Documentation

- A. General
- B. Model
- C. User Manual
- D. Computer Program

II. Software

- A. Verification
- B. User Friendliness
- C. General

III. Validation

- A. Technical Validity
- B. Operational Validity
- C. Data Validity

IV. Overall

- A. General
- B. Specific Recommendation

in general. The specifics of the overall recommendation is contained as a criterion in the category of overall and will be discussed in the next section.

Prior to presenting the specifics of the criteria, a simplified version of the modelling process [11, 12] will be presented to ensure a clear understanding of how certain criteria suggested for inclusion in this assessment procedure relate to the modelling process. Consider the basic elements of problem entity, conceptual model, and computerized model, and their relationships as depicted in Figure 1. The problem entity is the system (real or proposed), idea, situation, policy or phenomena under analysis; the conceptual model is a mathematical/logical/verbal representation of the problem entity appropriate for a particular study, and the computerized model is the conceptual model implemented on a computer. The conceptual model is developed through an analysis and modelling phase, the computerized model is developed through a computer programming and implementation phase, and inference about the problem entity are obtained by conducting computer experiments on the computerized model in the experimentation phase.

In this paper we will view the model verification and validation steps in the modelling process as consisting of conceptual model validation, computerized model verification, and operational validity as depicted in Figure 1. We define conceptual model validation as determining that the theories and assumptions underlying the conceptual model are correct, the modelling approach selected

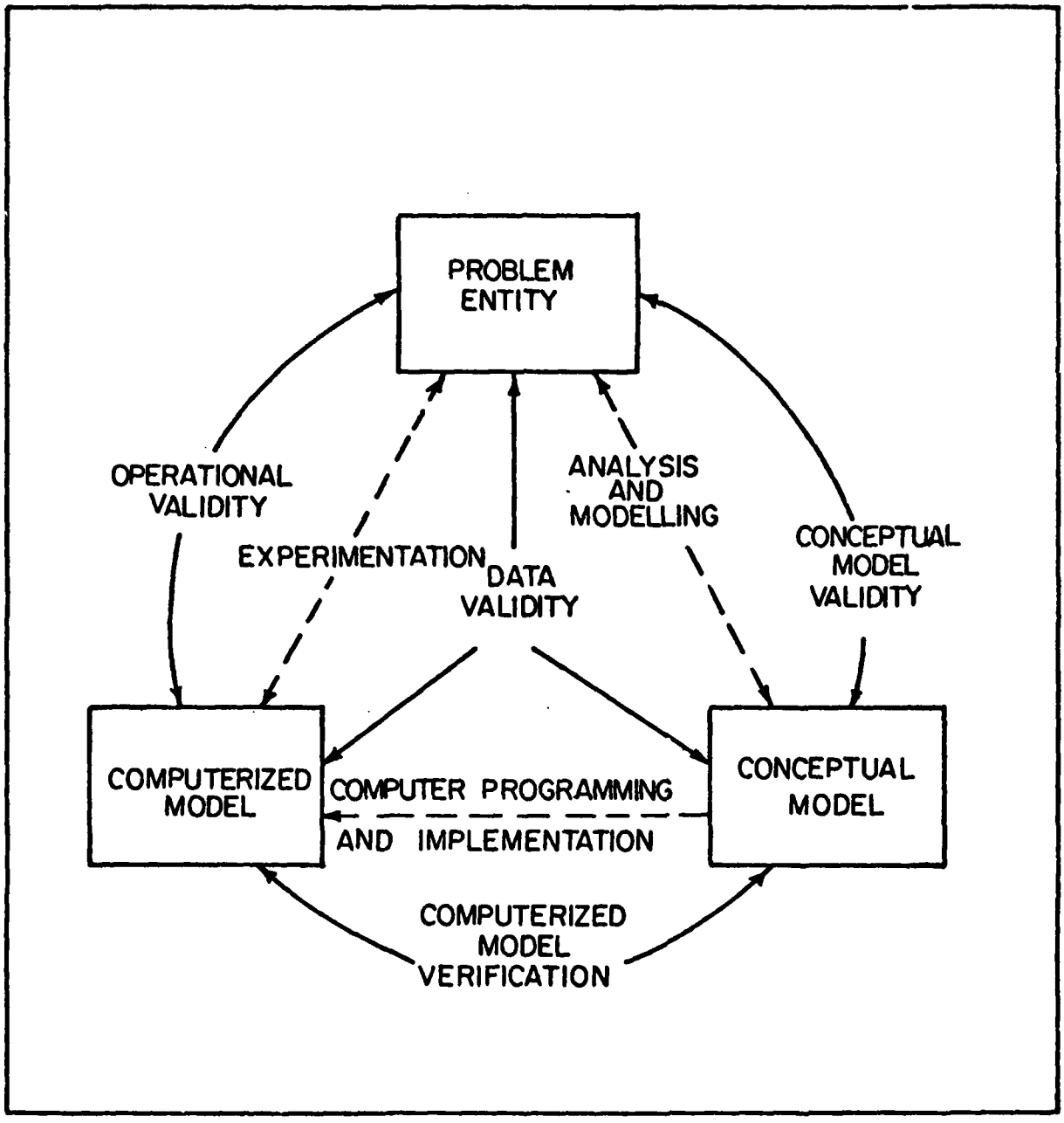


FIGURE 1. A Simplified Version of the Modelling Process.

is appropriate, and the model representation of the problem entity is "reasonable" for the intended use of the model. Computerized model verification is defined as ensuring that the computer programming and implementation of the conceptual model is correct. Operational validity is determining that the pertinent characteristics of the model adequately represent the problem entity for a specific use of the model. Data validity, also shown in Figure 1, is determining that the data necessary for model building and testing are adequate and correct.

ASSESSMENT CRITERIA

In this section we discuss the set of criteria suggested for use in assessing general purpose computerized models and computer-based modelling tools. The criteria will be presented by the categories of Documentation, Software, Validation, and Overall. The definition of each category and each criterion will be given as they are presented. The assessor will make a subjective rating between 0 and 10, with 0 being the lowest and 10 the highest, on how well each criterion specified by the assessment sponsor is satisfied. Some criteria require the assessment sponsor to supply certain specific information when they specify their use.

Specific details are not given in this paper on how an assessor is to make the subjective rating for each criterion. The assessment sponsor should usually specify to an assessor either how to make each subjective rating or how to use the subjective rating scale. Examples of how to use the subjective rating scale are (i) to use a threshold value. e.g., 4, which divides the scale into unsatisfactory and satisfactory range, or (ii) to divide the scale into groups such as 0-1 are unsatisfactory, 2-3 are poor, 4-6 are fair, 7-8 are good, and 9-10 are excellent.

DOCUMENTATION

Documentation is the written, or otherwise recorded, information or material presented with the computerized model or computer-based modelling tool and is usually presented as a set of documents. This author believes that for most assessments the documentation should be divided into different classifications and an evaluation

be made for each classification. The classification that this author believes appropriate for most assessments are General, Model or Modelling tool, User Manual, and Computer Program.

The assessment sponsor should specify for each document classification whom they should be written for, e.g., high management, technical users, non-technical users, or computer programmers, and the evaluation made with respect to the users specified. The assessment sponsor may specify, if appropriate, more than one type of user for each document classification. If more than one type of user is specified for a given classification, the assessment sponsor must specify whether they want a separate evaluation for each type of user, an overall evaluation for all users specified, or both, for that documentation classification. For each document evaluation specified by the assessment sponsor, the assessor should make an evaluation and give a subjective rating between 0 and 10 on how well the documentation satisfies the type(s) of users specified. The specifics of each document classification or criterion will be discussed next.

1. General Documentation: The general documentation should contain general information on the computerized model or computer-based modelling tool and a description of the documents for that model or modelling tool. This documentation would normally include as a minimum the title of the model or modelling tool, the model's or modelling tool's purpose, model sponsor, model developers, time frame of development, man-power or dollar effort expended in its development, type of model or modelling tool (e.g., interactive simulation model), major model assumptions, model data requirements,

availability, computer language used, computer system requirements, a description of the verification and validation performed, any major use(s) of it, and a description of the documentation provided.

2. Model or Modelling Tool Documentation: The model or modelling tool documentation should contain all of the specifics of the computerized model or computer-based modelling tool and should be detailed in nature. Generally, this documentation would include as a minimum, the type of model or modelling tool (e.g., queueing model or simulation), a detailed description of the purpose of the model or modelling tool, all underlying assumptions, the theories used, type of solution used, a detailed description of the model or models used (e.g., model flowchart for a simulation model or equations of an analytic model), data used in developing the model or modelling tool, rationale for the modelling approach used, type of uses the model or modelling tool is designed for, detailed information on verification and validation, type of model accuracy expected with range of applicability, data required to use the model or modelling tool, computer language used, mode of model or modelling tool usage (e.g., interactive), computer system requirements (e.g., storage requirement, computation time, and computer equipment), and adequate references.

3. User Manual: The user manual should contain all the information required for the specified type of users to be able to use the computerized model or computer-based modelling tool after

it has been implemented on a computer. This documentation would typically include, as a minimum, how to use the model or modelling tool in its various modes for different types of applications, the human-computer interfaces for its various modes (e.g., batch with the appropriate card statements), the type of data required and how to put it into the computer system, how various studies can be made using the software (e.g., sensitivity studies), how to use any analysis tools provided (e.g., statistical analysis), how to use any report generation capability, and specific examples that a user can reproduce.

4. Computer Program Documentation: The computer program documentation contains all of the documentation on the computer program and how to implement it on a computer system. This documentation would normally include, as a minimum, program flowcharts of various levels (macro to detail), documentation of the computer code, identification of computer code that is computer machine dependent, the type of computer language used, the availability of the program, specific verification procedures performed, how to implement the computer program on a computer system, and the computer system requirements such as storage and equipment required.

SOFTWARE

The category Software is concerned with the computerized model or computer-based modelling tool software. This author believes that the three criteria of Verification, User Friendliness, and General

are appropriate for most assessments. For each software criterion specified by the assessment sponsor, the assessor is to make a subjective rating between 0 and 10. The specifics of each criterion are discussed below.

1. Verification: Verification is determining that the conceptual model was programmed and implemented in software correctly. This is ensuring that the step called computerized model verification in the modelling process presented above was performed correctly. Verification assessment can be conducted at various indepth levels, at least in theory; but this author believes that most assessments would be performed only at one of three levels. One level is to only review the verification performed by the model developers; a second level is to review the verification performed by the model developers and to perform dynamic testing (analysis) [12], i.e., run the model under various conditions or use the modelling tool for different cases; and a third level is to review the verification performed by the model developers and to conduct indepth verification studies, e.g., performing static and dynamic analysis (testing) or reprogramming critical parts of the computer program. The assessment sponsor should specify the level of verification assessment desired, and any specific verification techniques that must, as a minimum, be performed by the assessor.

2. User Friendliness: User friendliness is determining how easy it is to use the computerized model or computer-based modelling tool software. This considers evaluating at least the following:

ease of understanding input requirements; ease of putting in input, e.g., manually using cards or terminal, using a data base, or using some type of computer interface; ease of using interaction media, e.g., batch or terminal; ease of understanding output; ease of obtaining desired and necessary output in desired form and formats; ease of generating reports; and overall ease of using the software.

3. General: The software general criterion is to evaluate all of the aspects of the software except for verification and user friendliness. This includes, as a minimum, the storage requirement of the software, the computational efficiency of the software, the computer system requirements for the software, the portability of the software, maintainability of the software, ease of software implementation, computer language(s) used, and whether good programming techniques (e.g., structured programming and program modularity) were used. The assessor would make the subjective rating for this criterion considering all of these factors.

VALIDATION

Validation consists of ensuring (1) that the general purpose computerized model or computer-based modelling tool is technically correct and can provide an adequate agreement between the type of systems or problem entities for its intended use, (2) that for each specific application of the model or modelling tool, an adequate representation of the model or modelling tool was obtained, and (3)

that the data used are correct and adequate. Because this assessment criteria are for general purpose models and modelling tools, the criteria are divided into two groups: General and Specific. The General group contains only one criteria and is called Technical Validity. The Specific group contains two criteria, Operational Validity and Data Validity, and they are used for specific applications of the model or modelling tool. The assessor makes a subjective rating for each of the two specific criteria for each application the assessment sponsor specifies to be assessed. The assessment sponsor must specify what type of validation is to be performed, e.g., only technical validity, technical validity and various specific assessments, or only a specific assessment. Each of the criterion will be discussed next.

1. Technical Validity: Technical validity is determining that the conceptual model is technically correct, i.e., ensuring the validity of the conceptual model, and that the computerized model or modelling tool can provide an adequate agreement between the type of systems or problem entities for its intended use. This usually includes, as a minimum, determining that the data and information used to develop the conceptual model are correct and adequate; any data transformations, e.g., aggregation, used are correct; the modelling approach selected is appropriate; the assumptions, theories, and hypotheses underlying the conceptual model are correct; the numerical computations made in the computerized model have sufficient accuracy; and the computerized model or

modelling tool can provide the desired accuracy with respect to the type of systems and problem entities that its use is intended for. It is expected that the assessor will exercise (run) the computerized model or computer-based modelling tool and use various validation techniques [11, 12, 14] in this evaluation, e.g., use face validity, perform sensitivity analysis, and make comparisons of model output to known results, e.g., using known results from simple analytical models, if possible. The assessor will, upon completion of the evaluation, make a subjective rating for technical validity.

2. Operational Validity: Operational validity is determining whether or not the computerized model or computer-based modelling tool in a specific application is providing an adequate representation of that system or problem entity for its intended use. In a specific application, parameter values will have to be entered into the computerized model or computer-based modelling tool to have that model or modelling tool represent that specific application. Thus, an assessor must ensure that the parameter values selected are correct.

Various levels or depths of operational validity assessment are possible and what is desired by the assessment sponsor should be specified. This author believes that the level of assessment desired would usually either be (1) to review what the model developers have performed, (2) to review and replicate what the model developers have performed, or (3) to perform an indepth operational validity study using the appropriate validation techniques [11, 12, 14]. Up-

on the completion of this evaluation, the assessors will make subjective rating.

3. Data Validity: Data validity is concerned with determining whether the data used in a specific application of the computerized model or computer-based modelling tool are correct. This consists of evaluating (1) the data used to obtain the values of the model or modelling tool parameters and (2) the data used to perform operational validity, e.g., input-output comparisons for various experimental frames [12, 16].

OVERALL

The Overall category consists of making an overall subjective rating of the computerized model or computer-based modelling tool and making a specific recommendation regarding its use.

1. General: The general criteria is used to make a subjective rating of the computerized model or computer-based modelling tool considering all of the previous subjective ratings and any additional factors that the assessor believes appropriate.

2. Specific Recommendation: The specific recommendation is for the assessor to make a specific recommendation to the assessment sponsor regarding the computerized model or computer-based modelling tool. The assessment sponsor should provide a list of alternatives for the assessor to select from. A typical list might be:

(1) Should not be used by any one;

(2) Should only be used by highly qualified technical analysts and with caution;

(3) Should only be used by highly qualified technical analysts;

(4) Can be used by any analyst but with caution;

(5) Can be used by any analyst;

(6) Can be used by any nontechnical individual but with caution;

(7) Can be used by any nontechnical individual;

(8) Can be used by anyone and is an outstanding model or modelling tool.

CONCLUSIONS

In this paper, a flexible and adaptive assessment procedure and a set of assessment criteria were presented for use in having assessments conducted of general purpose computerized models and computer-based modelling tools. Basically, the procedure requires the use of a third party to make subjective judgements on a set of criteria specified by the assessment sponsor and to make an overall recommendation on the model or modelling tool use. The set of criteria presented may also be useful in helping to specify to model developers what is expected of them by model sponsors.

REFERENCES

[1] Gass, S.I., "Evaluation of Complex Models," Computers and Operations Research, Vol. 4, pp. 27-35, March 1977.

[2] Gass, S.I., "A Procedure for the Evaluation of Complex Models," Proc. of the First International Conference on Mathematical Models, 1977.

[3] Gass, S.I., "Validation and Assessment Issues of Energy Models," NATO Advanced Research Institute on "Application of Systems Science to Energy Policy Planning," November 1979.

[4] Gass, S.I. (editor), Validation and Assessment Issues of Energy Models, National Bureau of Standards Publication 569, February 1980 (Conference held January 1979).

[5] Gass, S.I., et al., "Guidelines for Model Evaluation: An Abridged Version of the U.S. General Office Exposure Draft," Operations Research, Vol. 28, No. 2, March-April 1980.

[6] Gruhl, J. and D.O. Wood, "Independent Assessment of Complex Models," Proc. of the Workshop on Validation of Computer-based Mathematical Models in Energy Related Research and Development, edited by Hoffman, et al., October 1979.

[7] Hoffman, A.A.J., et al. (editors), Proceedings of the Workshop on Validation of Computer-based Mathematical Models in Energy Related Research and Development, October 1979 (held June 1978, Fort Worth, Texas).

[8] House, P.W., et al., Large-Scale Models for Policy Evaluation, Wiley Interscience, 1977.

[9] Kresge, D., "An Approach to Independent Model Assessment," Validation and Assessment Issues of Energy Models, Gass, S.I. (editor), National Bureau of Standards Publication 569, February 1980.

[10] Oren, T.I., "Concepts and Criteria to Assess Acceptibility of Simulation Studies: A Frame of Reference," forthcoming in Communications of ACM.

[11] Sargent, R.G., "Validation of Simulation Models," Proc. of 1979 Winter Simulation Conference, edited by H.J. Highland, et al.

[12] Sargent, R.G., "Verification and Validation of Simulation Models," forthcoming in Proc. of SIMULATION '80, edited by F. E. Cellier.

[13] Schellenberg, R.E., "Criteria for Assessing Model Validity for Managerial Purposes," Decision Science, Vol. 5, No. 4, October 1974.

[14] Shannon, R.E., System Simulation: The Art and Science, Prentice-Hall, Inc., 1975.

[15] U.S. General Accounting Office, "Guidelines for Model Evaluation," PAD-79-17, U.S. G.A.O., January 1979.

[16] Zeigler, B.P., Theory of Modelling and Simulation, John Wiley and Sons, Inc., 1976.

LME